

Intentions and assignments

Emiliano Lorini¹, Mehdi Dastani², Hans van Ditmarsch³,
Andreas Herzig¹, John-Jules Meyer²

¹ Université de Toulouse, IRIT-CNRS, France

² Utrecht University, The Netherlands

³ University of Sevilla, Spain

Abstract. The aim of this work is propose a logical approach to intention dynamics based on the notion of *assignment* [7, 3]. The function of an assignment is to associate the truth value of a certain formula φ to a propositional atom p . We combine a static modal logic of belief and choice with three kinds of dynamic modalities and corresponding three kinds of assignments: assignments operating on an agent's beliefs, assignments operating on the agent's choices and assignments operating on the objective world. An agent's intention is defined in our approach as the agent's choice to perform a given action and two basic operations on intentions called intention generation and intention reconsideration are defined as specific kinds of assignments on choices.

1 Introduction

Since the seminal work of Cohen & Levesque [6] aimed at implementing Bratman's philosophical theory of intention [5], many formal logics for reasoning about intentions, their dynamics and their relationships with beliefs have been developed (see, e.g., [14, 16, 17, 13, 19, 12]). Most of them are frameworks based on a blend of dynamic logic with doxastic logic, enriched with modal operators for motivational attitudes such as preferences, goals and intentions. But, although logical analysis of intention dynamics are available in the literature, the issue of a *formal semantics* for the dynamics of intentions has received much less attention. Indeed, all previous approaches are mostly interested in characterizing in the object language the epistemic conditions under which an agent's intention persists over time and the epistemic conditions under which an agent's intention is generated, but they do not provide a semantic characterization of the process of generating an intention and of the process of reconsidering an intention.

The aim of this work is to shed light on this unexplored area by proposing a formal semantics of intention dynamics based on the notion of *assignment*. The function of an assignment is to associate the truth value of a certain formula φ to a propositional atom p . We combine a static modal logic including modal operators for belief and choice with three kinds of dynamic modalities and corresponding three kinds of assignments: assignments operating on an agent's beliefs, assignments operating on the agent's choices and assignments operating on the objective world. An agent's intention is defined in our approach as the agent's choice to perform a given action and two basic operations on intentions called *intention generation* and *intention reconsideration* are defined as specific kinds of assignments on choices.

Assignments were studied before in the literature on modal logic for information dynamics. However they were only applied to the dynamics of belief and knowledge [7, 3], and there is still no application of this notion to the theory of intention. In this paper we show that assignments are well-suited to model intention dynamics. Indeed, assignments capture the *locality* of intention dynamics better than other operations like announcements [8] and upgrades [2, 11], where locality means that the process of re-considering (or generating) an agent’s intention does not affect the other intentions of the agent.

The rest of the paper is organized as follows. The first part (Section 2) introduces a static logic of belief, choice and intention. In the second part (Section 3) we move from a static perspective on agents’ attitudes to a dynamic perspective. We first present the syntax and semantics of three kinds of assignments on beliefs, on choices and on the objective world. Then in Section 4, we analyze the notion of *executability preconditions* for assignments. We devote special attention to *executability preconditions* of assignments which are responsible for the generation (resp. reconsideration) of an intention. In Section 5 we compare our approach with existing logical approaches to belief and preference dynamics.

2 A modal logic of beliefs, choices and intentions

We introduce a modal logic called **L** which supports reasoning about three different kinds of mental attitudes: beliefs, choices (or chosen goals), and intentions.

2.1 Syntax

Let $ATM^{Fact} = \{f_1, f_2, \dots\}$ be a nonempty finite set of atoms denoting facts (or state of affairs), and let $ATM^{Act} = \{\alpha, \beta, \dots\}$ be a nonempty finite set of atoms denoting actions. The atom α is meant to stand for ‘the agent performs a certain action α ’. We also have special atoms of type $good_\alpha$ expressing that ‘performing action α is good for the agent’. ATM^{Good} is the corresponding set, that is, $ATM^{Good} = \{good_\alpha | \alpha \in ATM^{Act}\}$. We define $ATM = ATM^{Fact} \cup ATM^{Act} \cup ATM^{Good}$ to be the set of atomic formulas. We note p, q, \dots the elements in ATM .

The language \mathcal{L} of the logic **L** is the set of formulas defined by the following BNF:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [B]\varphi \mid [C]\varphi$$

where p ranges over ATM . The other Boolean constructions \top , \perp , \vee , \rightarrow and \leftrightarrow are defined from \neg and \wedge in the standard way.

The two modal operators of our logic have the following reading: $[B]\varphi$ means ‘the agent believes φ ’ and $[C]\varphi$ means ‘the agent has chosen φ ’ (or ‘the agent wants φ to be true’). Operators $[C]$ are used to denote the agent’s choices, that is, the state of affairs that the agent has decided to pursue. Similar operators have been studied in [6, 12, 14].

The following abbreviation will also be convenient for every $\alpha \in ATM^{Act}$:

$$I(\alpha) \stackrel{\text{def}}{=} [C]\alpha.$$

$I(\alpha)$ is meant to stand for ‘the agent intends to do action α ’.

2.2 Semantics

Models of the logic **L** (**L**-models) are tuples $F = \langle W, \mathcal{B}, \mathcal{C}, \mathcal{V} \rangle$ defined as follows:

- W is a nonempty set of possible worlds or states;
- $\mathcal{B} \subseteq W \times W$ is a serial, transitive and Euclidean accessibility relation for belief;
- $\mathcal{C} \subseteq W \times W$ is a serial, transitive and Euclidean accessibility relation for choice;
- $\mathcal{V} : ATM \rightarrow 2^W$ is a valuation function.

Accessibility relations on W can be viewed as functions from W to 2^W . Therefore, we write $\mathcal{B}(w) = \{v \mid (w, v) \in \mathcal{B}\}$ and $\mathcal{C}(w) = \{v \mid (w, v) \in \mathcal{C}\}$. $\mathcal{B}(w)$ is the set of worlds that are compatible with the agent's beliefs at w (or belief accessible worlds at w), $\mathcal{C}(w)$ is the set of worlds that are compatible with the agent's choices at w (or choice accessible worlds at w).

The accessibility relations \mathcal{B} and \mathcal{C} satisfy the following additional constraints for every $w \in W$:

- S1 if $v \in \mathcal{B}(w)$ then $\mathcal{C}(v) = \mathcal{C}(w)$;
- S2 if $v \in \mathcal{C}(w)$ and $u \in \mathcal{B}(v)$ then $u \in \mathcal{C}(w)$;
- S3 if $v \in \mathcal{C}(w)$ then $v \in \mathcal{B}(v)$.

Constraint S1 expresses that the agent's choices are positively and negatively introspective (i.e. if v is compatible with the agent's beliefs at w then the set of worlds which are compatible with the agent's choices at w is identical to the set of worlds which are compatible with the agent's choices at v). According to constraint S2, the agent always chooses the states that he considers possible from the states that he chooses. According to constraint S3, if at state w the agent chooses state v then at v the agent considers v a possible state. In other words, the agent always chooses that the current state belongs to the set of states that he considers possible.

Given a model M , a world w and a formula φ , we write $M, w \models \varphi$ to mean that φ is true at world w in M . The rules defining the truth conditions of formulas are just standard for atomic formulas, negation and disjunction. The following are the remaining truth conditions for $[B]\varphi$, $[C]\varphi$:

- $M, w \models [B]\varphi$ iff $M, v \models \varphi$ for all w' such that $v \in \mathcal{B}(w)$;
- $M, w \models [C]\varphi$ iff $M, v \models \varphi$ for all w' such that $v \in \mathcal{C}(w)$.

We write $\models_{\mathbf{L}} \varphi$ if φ is *valid* (i.e. φ is true in all **L**-models). We say that φ is *satisfiable* if $\neg\varphi$ is not valid.

2.3 Axiomatization

Fig. 1 contains the axiomatization of the logic **L**. We adopt a standard KD45 logic for beliefs [9] and a standard KD45 logic for choices [6]. Thus, we have positive and negative introspection for beliefs (Axioms 4 and 5 for $[B]$), and we assume that an agent cannot have inconsistent beliefs (Axiom D for $[B]$). We assume that if the agent chooses something then he chooses to choose it and if the agent does not choose something then he chooses not to choose it (Axioms 4 and 5 for $[C]$), and we assume that the

agent cannot have inconsistent choices (Axiom D for $[C]$). We have negative and positive introspection for choices (Axioms **PI** $\text{Intr}_{[C]}$ and **NI** $\text{Intr}_{[C]}$). Moreover, we suppose that if the agent chooses φ then he chooses to believe φ (Axiom **AchieveAware**). In other words, if the agent wants that φ will be true then he wants to achieve a state in which he believes that φ is true (i.e. he wants to achieve φ knowingly). Finally, we suppose that the agent always wants that if φ is true then he does not believe $\neg\varphi$ (Axiom **NotIncorrectBel**). In other words, the agent always wants not to have incorrect beliefs.

(PC)	All principles of classical propositional calculus
(KD45) $_{[B]}$	All principles of modal logic KD45 for $[B]$
(KD45) $_{[C]}$	All principles of modal logic KD45 for $[C]$
(PI) $\text{Intr}_{[C]}$	$[C]\varphi \rightarrow [B][C]\varphi$
(NI) $\text{Intr}_{[C]}$	$\neg[C]\varphi \rightarrow [B]\neg[C]\varphi$
(AchieveAware)	$[C]\varphi \rightarrow [C][B]\varphi$
(NotIncorrectBel)	$[C](\varphi \rightarrow \neg[B]\neg\varphi)$

Fig. 1. Axiomatization of **L**

We call **L** the logic axiomatized by the principles given in Fig. 1. We write $\vdash_{\mathbf{L}} \varphi$ if φ is a **L**-theorem. For instance, the following theorem is provable by Axiom **NotIncorrectBel**, Axiom K and necessitation rule for $[C]$:

$$\vdash_{\mathbf{L}} [C][B]\varphi \rightarrow [C]\varphi.$$

The theorem just says that if the agent wants to acquire a certain belief then he wants the content of this belief to be true. Therefore, our logic does not allow self-deception. For example, it excludes the situation of a person who wants to believe that God exists in order to feel better when thinking about the afterlife (i.e. $[C][B]GodExists$) and, at the same time, she does not want that God exists (i.e. $\neg[C]GodExists$).

It is to be noted that $[B]\varphi \wedge [C]\neg\varphi$ are satisfiable in our logic. For example, our logic allows the situation of a person who smokes and believes this (i.e. $[B]smoke$) and, she decides to stop smoking (i.e. $[C]\neg smoke$).

Theorem 1. *The logic **L** is completely axiomatized by the principles in Fig. 1.*

Proof. It is a routine task to check that the axioms of the logic **L** correspond one-to-one to their semantic counterparts on the models. In particular, Axioms D, 4 and 5 for $[B]$ correspond to the seriality, transitivity and Euclideanity of the accessibility relation \mathcal{B} . Axioms D, 4 and 5 for $[C]$ correspond to the seriality, transitivity and Euclideanity of \mathcal{C} . Axioms **PI** $\text{Intr}_{[C]}$ and **NI** $\text{Intr}_{[C]}$ together correspond to the constraint S1. Finally, Axiom **AchieveAware** corresponds to S2 and Axiom **NotIncorrectBel** corresponds to S3. It is routine, too, to check that all of our axioms are in the Sahlqvist class. This means that the axioms are all expressible as first-order conditions on models and that they are complete with respect to the defined model classes, cf. [4, Th. 2.42]. \square

3 From static to dynamic mental states

In this section we extend the logic \mathbf{L} of Section 2 by modal operators for objective world change and mental attitude change. We distinguish two kinds of mental attitude change: belief change and choice change. We call \mathbf{L}^{dyn} the extended logic.

3.1 Syntax

Atomic events for world change (or *atomic world assignments*) are of the form $p \overset{W}{\rightsquigarrow} \varphi$ whereas atomic events for belief change (or *atomic belief assignments*) and for choice change (or *atomic choice assignments*) are of the form $p \overset{B}{\rightsquigarrow} \varphi$ and $p \overset{C}{\rightsquigarrow} \varphi$: $p \overset{B}{\rightsquigarrow} \varphi$ is the event ‘the truth value of φ is assigned to p in the agent’s beliefs’; $p \overset{C}{\rightsquigarrow} \varphi$ is the event ‘the truth value of φ is assigned to p in the agents’ choices’; $p \overset{W}{\rightsquigarrow} \varphi$ is the event ‘the truth value of φ is assigned to p in the objective world’.

We respectively note $BASG_B = \{p \overset{B}{\rightsquigarrow} \varphi | p \in ATM \text{ and } \varphi \in \mathcal{L}\}$, $BASG_C = \{p \overset{C}{\rightsquigarrow} \varphi | p \in ATM \text{ and } \varphi \in \mathcal{L}\}$ and $BASG_W = \{p \overset{W}{\rightsquigarrow} \varphi | p \in ATM \text{ and } \varphi \in \mathcal{L}\}$ the set of atomic belief assignments, the set of atomic choice assignments and the set of atomic world assignments. We define $EVT = BASG_B \cup BASG_C \cup BASG_W$ to be the set of atomic assignments.

The following two types of atomic choice assignments characterize two basic operations on an agent’s intentions:

$$\begin{aligned} gen(\alpha) &\stackrel{\text{def}}{=} \alpha \overset{C}{\rightsquigarrow} \top; \\ rec(\alpha) &\stackrel{\text{def}}{=} \alpha \overset{C}{\rightsquigarrow} \perp. \end{aligned}$$

The event $gen(\alpha)$ is the agent’s mental operation of *generating* the intention to do action α , whereas the event $rec(\alpha)$ is the agent’s mental operation of *reconsidering* (or *erasing*) his intention to perform action α .⁴

Complex assignments are defined as *partial functions* from ATM to \mathcal{L} . We distinguish three kinds of complex assignments: *complex belief assignments*, *complex choice assignments* and *complex world assignments*. We note $\sigma_B, \sigma'_B, \dots$ the complex belief assignments, $\sigma_C, \sigma'_C, \dots$ the complex choice assignments and $\sigma_W, \sigma'_W, \dots$ the complex world assignments. Moreover, we respectively note $CASG_B$, $CASG_C$ and $CASG_W$ the set of all complex belief assignments, the set of all complex choice assignments and the set of all complex world assignments. Given a complex belief assignment σ_B , $D(\sigma_B)$ is the *domain* of σ_B and $C(\sigma_B)$ is its *codomain*. Similarly, $D(\sigma_C)$ (resp. $D(\sigma_W)$) is the domain of the complex choice (resp. world) assignment σ_C (resp. σ_W) and $C(\sigma_C)$ (resp. $C(\sigma_W)$) is its codomain. We extend partial functions to total functions by stipulating that when $p \notin D(\sigma_B)$ then $\sigma_B(p) = p$. Similarly, we stipulate that $\sigma_C(p) = p$ when $p \notin D(\sigma_C)$, and $\sigma_W(p) = p$ when $p \notin D(\sigma_W)$.

For every complex belief assignment σ_B , we define

$$s(\sigma_B) = \{p \overset{B}{\rightsquigarrow} \sigma_B(p) | p \in D(\sigma_B)\}$$

to be the corresponding set of atomic belief assignments. Similarly, for every complex

⁴ For some different approaches to intention generation and intention reconsideration, see e.g. [13, 10].

choice assignment σ_B and complex world assignments σ_W

$$s(\sigma_C) = \{p \overset{C}{\rightsquigarrow} \sigma_C(p) | p \in D(\sigma_C)\} \text{ and}$$

$$s(\sigma_W) = \{p \overset{W}{\rightsquigarrow} \sigma_W(p) | p \in D(\sigma_W)\}$$

are the corresponding sets of atomic choice and atomic world assignments.

The elements of ASG are all sets including a set of atomic belief assignments, a set of atomic choice assignments and a set of atomic world assignments, that is,

$$ASG = \{\{s(\sigma_B), s(\sigma_C), s(\sigma_W)\} | \sigma_B \in CASG_B, \sigma_C \in CASG_C, \sigma_W \in CASG_W\}.$$

The language \mathcal{L}^{dyn} of the logic L^{dyn} is defined by the following BNF:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [B]\varphi \mid [C]\varphi \mid [\Sigma:W]\psi \mid [\Sigma:B]\psi \mid [\Sigma:C]\psi$$

where p ranges over ATM and Σ ranges over ASG .

The formula $[\Sigma:W]\psi$ is meant to stand for: ψ holds in the objective world after the occurrence of the event Σ . The formula $[\Sigma:B]\psi$ is meant to stand for: ψ holds in the context of the agent's beliefs after the occurrence of the event Σ . The formula $[\Sigma:C]\psi$ is meant to stand for: ψ holds in the context of the agent's choices after the occurrence of the event Σ . The duals of the operators $[\Sigma:W]$, $[\Sigma:B]$ and $[\Sigma:C]$ are defined as usual: $\langle \Sigma:W \rangle \psi \stackrel{\text{def}}{=} \neg[\Sigma:W]\neg\psi$, $\langle \Sigma:B \rangle \psi \stackrel{\text{def}}{=} \neg[\Sigma:B]\neg\psi$, and $\langle \Sigma:C \rangle \psi \stackrel{\text{def}}{=} \neg[\Sigma:C]\neg\psi$.

3.2 Semantics

We introduce a function Pre from EVT to \mathcal{L} which returns the *executability preconditions* of every atomic belief assignment, of every atomic choice assignment and of every atomic world assignment. The function Pre is generalized to the events Σ in ASG in a straightforward manner. Suppose $\Sigma = \{s(\sigma_B), s(\sigma_C), s(\sigma_W)\}$. Then:

$$Pre(\Sigma) = \bigwedge_{p \overset{B}{\rightsquigarrow} \varphi \in s(\sigma_B)} Pre(p \overset{B}{\rightsquigarrow} \varphi) \wedge \bigwedge_{p \overset{C}{\rightsquigarrow} \varphi \in s(\sigma_C)} Pre(p \overset{C}{\rightsquigarrow} \varphi) \wedge \bigwedge_{p \overset{W}{\rightsquigarrow} \varphi \in s(\sigma_W)} Pre(p \overset{W}{\rightsquigarrow} \varphi).$$

Note that this formula is indeed in the language and is not infinitary. Indeed, the set of atoms ATM has been supposed to be finite.

For every $\Sigma \in ASG$, $Pre(\Sigma)$ denotes the executability preconditions of the event Σ , i.e. the conditions which are together necessary and sufficient to ensure that the event Σ will possibly *occur*.

Suppose that $\Sigma = \{s(\sigma_B), s(\sigma_C), s(\sigma_W)\}$. In order to give semantics to the operators $[\Sigma:W]$, $[\Sigma:B]$ and $[\Sigma:C]$ we define the model

$$M^\Sigma = \langle W^\Sigma, \mathcal{B}^\Sigma, \mathcal{C}^\Sigma, \mathcal{V}^\Sigma \rangle$$

which results from the occurrence of the event Σ in the model M . The elements of M^Σ

are defined as follows:

$$\begin{aligned}
W^\Sigma &= \{w_W | w \in W \text{ and } M, w \models \text{Pre}(\Sigma)\} \cup \\
&\quad \{w_B | w \in W \text{ and } M, w \models \text{Pre}(\Sigma)\} \cup \\
&\quad \{w_C | w \in W \text{ and } M, w \models \text{Pre}(\Sigma)\}; \\
\mathcal{B}^\Sigma &= \{(w_W, v_B) | v, w \in W \text{ and } (w, v) \in \mathcal{B}\} \cup \\
&\quad \{(w_B, v_B) | v, w \in W \text{ and } (w, v) \in \mathcal{B}\} \cup \\
&\quad \{(w_C, v_C) | v, w \in W \text{ and } (w, v) \in \mathcal{B}\}; \\
\mathcal{C}^\Sigma &= \{(w_W, v_C) | v, w \in W \text{ and } (w, v) \in \mathcal{C}\} \cup \\
&\quad \{(w_B, v_C) | v, w \in W \text{ and } (w, v) \in \mathcal{C}\} \cup \\
&\quad \{(w_C, v_C) | v, w \in W \text{ and } (w, v) \in \mathcal{C}\}; \\
\mathcal{V}^\Sigma(p) &= \{w_W | w \in W \text{ and } M, w \models \sigma_W(p)\} \cup \\
&\quad \{w_B | w \in W \text{ and } M, w \models \sigma_B(p)\} \cup \\
&\quad \{w_C | w \in W \text{ and } M, w \models \sigma_C(p)\}.
\end{aligned}$$

M^Σ is obtained by creating three copies of each state of the original model M (a copy for the objective world, a copy for belief, a copy for choice), and by restricting the original model to the set of states in which the executability preconditions of Σ hold. Moreover, for every atom p , the effect of a model update by Σ is to assign the truth value of $\sigma_B(p)$ to the atom p in all belief copies of the original states, to assign the truth value of $\sigma_C(p)$ to the atom p in all choice copies, and to assign the truth value of $\sigma_W(p)$ to the atom p in all world copies.

For every world copy w_W , at w_W the agent considers possible all belief copies of those states that he considered possible before the event Σ , and he chooses all choice copies of those states that he chose before the event Σ .

For every belief copy w_B , at w_B the agent considers possible all belief copies of those states that he considered possible before the event Σ , and he chooses all choice copies of those states that he chose before the event Σ .

For every choice copy w_C , at w_C the agent considers possible all choice copies of those states that he considered possible before the event Σ , and he chooses all choice copies of those states that he chose before the event Σ .

This construction of the updated model M^Σ ensures that the agent is aware that his choices have been changed accordingly so that the properties of positive and introspection over the agent's choices (constraint S1) are preserved after the occurrence of the event Σ . Moreover, it ensures that the agent chooses that his beliefs change as his choices change so that the constraints S2 and S3 are preserved after the occurrence of the event Σ . On the contrary, the operation of world change is independent of the operations of belief and choice change, and the operations of belief and choice change are independent of the operation of world change.

Theorem 2. *If M is an L-model then M^Σ is an L-model.*

Proof. It is just trivial to prove that our operation of model update preserves the seriality, transitivity and Euclideanity of the accessibility relations \mathcal{B} and \mathcal{C} .

We just prove that the constraints S1, S2 and S3 are also preserved.

We start with S1. Assume $v_B \in \mathcal{B}^\Sigma(w_W)$ and $u_C \in \mathcal{C}^\Sigma(v_B)$. It follows that $v \in \mathcal{B}(w)$ and $u \in \mathcal{C}(v)$ and $M, w \models \text{Pre}(\Sigma)$ and $M, u \models \text{Pre}(\Sigma)$. Then, by constraint S1, we have $u \in \mathcal{C}(w)$. Therefore, $u_C \in \mathcal{C}^\Sigma(w_W)$. In a similar way we can prove that if $v_B \in \mathcal{B}^\Sigma(w_B)$ and $u_C \in \mathcal{C}^\Sigma(v_B)$ then $u_C \in \mathcal{C}^\Sigma(w_B)$, and if $v_C \in \mathcal{B}^\Sigma(w_C)$ and $u_C \in \mathcal{C}^\Sigma(v_C)$ then $u_C \in \mathcal{C}^\Sigma(w_C)$.

Now, assume $v_C \in \mathcal{C}^\Sigma(w_W)$ and $u_B \in \mathcal{B}^\Sigma(w_W)$. It follows that $v \in \mathcal{C}(w)$ and $u \in \mathcal{B}(w)$ and $M, v \models \text{Pre}(\Sigma)$ and $M, u \models \text{Pre}(\Sigma)$. Then, by constraint S1, we have $v \in \mathcal{C}(u)$. Therefore, $v_C \in \mathcal{C}^\Sigma(u_B)$. In a similar way we can prove that if $v_C \in \mathcal{C}^\Sigma(w_B)$ and $u_B \in \mathcal{B}^\Sigma(w_B)$ then $v_C \in \mathcal{C}^\Sigma(u_B)$, and if $v_C \in \mathcal{C}^\Sigma(w_C)$ and $u_C \in \mathcal{B}^\Sigma(w_C)$ then $v_C \in \mathcal{C}^\Sigma(u_C)$.

Let us prove S2. Assume $v_C \in \mathcal{C}^\Sigma(w_W)$ and $u_C \in \mathcal{B}^\Sigma(v_C)$. It follows that $v \in \mathcal{C}(w)$ and $u \in \mathcal{B}(v)$ and $M, w \models \text{Pre}(\Sigma)$ and $M, u \models \text{Pre}(\Sigma)$. Then, by constraint S2, we have $u \in \mathcal{C}(w)$. We can conclude that $u_C \in \mathcal{C}^\Sigma(w_W)$. In a similar way we can prove that if $v_C \in \mathcal{C}^\Sigma(w_B)$ and $u_C \in \mathcal{B}^\Sigma(v_C)$ then $u_C \in \mathcal{C}^\Sigma(w_B)$; and if $v_C \in \mathcal{C}^\Sigma(w_C)$ and $u_C \in \mathcal{B}^\Sigma(v_C)$ then $u_C \in \mathcal{C}^\Sigma(w_C)$.

Let us prove S3. Assume $v_C \in \mathcal{C}^\Sigma(w_W)$. It follows that $v \in \mathcal{C}(w)$ and $M, v \models \text{Pre}(\Sigma)$. Then, by constraint S3, we have that $v \in \mathcal{B}(v)$. Therefore, $v_C \in \mathcal{B}^\Sigma(v_C)$.

In a similar way we can prove that if $v_C \in \mathcal{C}^\Sigma(w_B)$ then $v_C \in \mathcal{B}^\Sigma(v_C)$ and if $v_C \in \mathcal{C}^\Sigma(w_C)$ then $v_C \in \mathcal{B}^\Sigma(v_C)$. \square

The truth conditions are those of Section 2 plus the following:

- $M, w \models [\Sigma:w]\psi$ iff, if $M, w \models \text{Pre}(\Sigma)$ then $M^\Sigma, w_W \models \psi$;
- $M, w \models [\Sigma:B]\psi$ iff, if $M, w \models \text{Pre}(\Sigma)$ then $M^\Sigma, w_B \models \psi$;
- $M, w \models [\Sigma:C]\psi$ iff, if $M, w \models \text{Pre}(\Sigma)$ then $M^\Sigma, w_C \models \psi$;

Note that $\langle \Sigma:w \rangle \top$ and $\langle \Sigma:B \rangle \top$ and $\langle \Sigma:C \rangle \top$ are individually equivalent to the executability preconditions of Σ (i.e. $\text{Pre}(\Sigma)$). Therefore $\langle \Sigma:w \rangle \top$, $\langle \Sigma:B \rangle \top$ and $\langle \Sigma:C \rangle \top$ should be respectively read ‘ Σ will possibly occur in the objective world’, ‘ Σ will possibly occur in the context of the agent’s beliefs’ and ‘ Σ will possibly occur in the context of the agent’s choices’.

In Section 4 we will specify in detail four general kinds of executability preconditions: the executability preconditions of the atomic world assignments $p \overset{W}{\rightsquigarrow} \perp$ and $p \overset{W}{\rightsquigarrow} \top$ (with $p \in \text{ATM}^{\text{Fact}}$); the executability preconditions of the atomic world assignment $\alpha \overset{W}{\rightsquigarrow} \top$ (with $\alpha \in \text{ATM}^{\text{Act}}$); the executability preconditions of the atomic choice assignment $\text{gen}(\alpha)$; the executability preconditions of the atomic choice assignment $\text{rec}(\alpha)$. The first are the executability preconditions of the event which consists in making true (resp. false) a certain objective fact p ; the second are executability preconditions of the intentional action α ; the third are the executability preconditions of the process of generating the intention to do α ; the fourth are the executability preconditions of the process reconsidering the intention to do α .

A discussion about the relationships between the present approach and the action/event model with assignments à la [7, 3] is given in Section 5.

3.3 Axiomatization

We have reduction axioms for the three operators $[\Sigma:w]$, $[\Sigma:B]$ and $[\Sigma:C]$.

Theorem 3. *Suppose $\Sigma = \{s(\sigma_B), s(\sigma_C), s(\sigma_W)\}$. Then, the following schemata are valid in \mathbf{L}^{dyn} :*

- R1a.** $[\Sigma:w]p \leftrightarrow (Pre(\Sigma) \rightarrow \sigma_W(p))$
- R1b.** $[\Sigma:B]p \leftrightarrow (Pre(\Sigma) \rightarrow \sigma_B(p))$
- R1c.** $[\Sigma:C]p \leftrightarrow (Pre(\Sigma) \rightarrow \sigma_C(p))$
- R2a.** $[\Sigma:w]\neg\varphi \leftrightarrow (Pre(\Sigma) \rightarrow \neg[\Sigma:w]\varphi)$
- R2b.** $[\Sigma:B]\neg\varphi \leftrightarrow (Pre(\Sigma) \rightarrow \neg[\Sigma:B]\varphi)$
- R2c.** $[\Sigma:C]\neg\varphi \leftrightarrow (Pre(\Sigma) \rightarrow \neg[\Sigma:C]\varphi)$
- R3a.** $[\Sigma:w](\varphi \wedge \psi) \leftrightarrow ([\Sigma:w]\varphi \wedge [\Sigma:w]\psi)$
- R3b.** $[\Sigma:B](\varphi \wedge \psi) \leftrightarrow ([\Sigma:B]\varphi \wedge [\Sigma:B]\psi)$
- R3c.** $[\Sigma:C](\varphi \wedge \psi) \leftrightarrow ([\Sigma:C]\varphi \wedge [\Sigma:C]\psi)$
- R4a.** $[\Sigma:w][B]\varphi \leftrightarrow (Pre(\Sigma) \rightarrow [B][\Sigma:B]\varphi)$
- R4b.** $[\Sigma:B][B]\varphi \leftrightarrow (Pre(\Sigma) \rightarrow [B][\Sigma:B]\varphi)$
- R4c.** $[\Sigma:C][B]\varphi \leftrightarrow (Pre(\Sigma) \rightarrow [B][\Sigma:C]\varphi)$
- R5a.** $[\Sigma:w][C]\varphi \leftrightarrow (Pre(\Sigma) \rightarrow [C][\Sigma:C]\varphi)$
- R5b.** $[\Sigma:B][C]\varphi \leftrightarrow (Pre(\Sigma) \rightarrow [C][\Sigma:C]\varphi)$
- R5c.** $[\Sigma:C][C]\varphi \leftrightarrow (Pre(\Sigma) \rightarrow [C][\Sigma:C]\varphi)$

Proof. We just prove **R4a** as an example.

$M, w \models [\Sigma:w][B]\varphi$

IFF if $M, w \models Pre(\Sigma)$ then $M^\Sigma, w_W \models [B]\varphi$

IFF if $M, w \models Pre(\Sigma)$ then, if $v_B \in \mathcal{B}^\Sigma(w_W)$ then $M^\Sigma, v_B \models \varphi$

IFF if $M, w \models Pre(\Sigma)$ then, if $v \in \mathcal{B}(w)$ and $M, w \models Pre(\Sigma)$ and $M, v \models Pre(\Sigma)$ then $M^\Sigma, v_B \models \varphi$

IFF if $M, w \models Pre(\Sigma)$ then, if $v \in \mathcal{B}(w)$ and $M, v \models Pre(\Sigma)$ then $M^\Sigma, v_B \models \varphi$

IFF if $M, w \models Pre(\Sigma)$ then, if $v \in \mathcal{B}(w)$ then if $M, v \models Pre(\Sigma)$ then $M^\Sigma, v_B \models \varphi$

IFF if $M, w \models Pre(\Sigma)$ then, if $v \in \mathcal{B}(w)$ then $M, v \models [\Sigma:B]\varphi$

IFF if $M, w \models Pre(\Sigma)$ then $M, w \models [B][\Sigma:B]\varphi$

IFF if $M, w \models Pre(\Sigma) \rightarrow [B][\Sigma:B]\varphi$.

□

We call \mathbf{L}^{dyn} the logic axiomatized by the principles of the logic \mathbf{L} plus the axiom schemata of Theorem 3 and the rule of replacement of proved equivalence. We write $\vdash_{\mathbf{L}^{dyn}} \varphi$ if φ is a \mathbf{L}^{dyn} -theorem. The following are examples of \mathbf{L}^{dyn} -theorems about intention generation and intention reconsideration.

Proposition 1.

- (1a) $\vdash_{\mathbf{L}^{dyn}} [\{\emptyset, \{gen(\alpha)\}, \emptyset\}:w] \mathbf{I}(\alpha)$
- (1b) $\vdash_{\mathbf{L}^{dyn}} [\{\emptyset, \{rec(\alpha)\}, \emptyset\}:w] \neg \mathbf{I}(\alpha)$
- (1c) $\vdash_{\mathbf{L}^{dyn}} [\{\emptyset, \{gen(\alpha)\}, \emptyset\}:c] \alpha$
- (1d) $\vdash_{\mathbf{L}^{dyn}} [\{\emptyset, \{rec(\alpha)\}, \emptyset\}:c] \neg \alpha$
- (1e) $\vdash_{\mathbf{L}^{dyn}} \neg \mathbf{I}(\beta) \rightarrow [\{\emptyset, \{gen(\alpha)\}, \emptyset\}:w] \neg \mathbf{I}(\beta) \quad \text{if } \alpha \neq \beta$
- (1f) $\vdash_{\mathbf{L}^{dyn}} \mathbf{I}(\beta) \rightarrow [\{\emptyset, \{rec(\alpha)\}, \emptyset\}:w] \mathbf{I}(\beta) \quad \text{if } \alpha \neq \beta$

Proof. We prove Theorem (1a) as an example. $[\{\emptyset, \{gen(\alpha)\}, \emptyset\}:w] \mathbf{I}(\alpha)$ is equivalent to $Pre(\Sigma) \rightarrow [C][\{\emptyset, \{gen(\alpha)\}, \emptyset\}:c] \alpha$ (by reduction axiom **R5a**. and rule of replacement of proved equivalences). The latter is equivalent to $Pre(\Sigma) \rightarrow [C](Pre(\Sigma) \rightarrow \top)$ (by reduction axiom **R1c**. and rule of replacement of proved equivalences) which in turn is equivalent to \top . \square

According to Theorem 1a, after generating the intention to do α , the agent intends to do α in the objective world. According to Theorem 1b, after reconsidering the intention to do α , the agent does not intend to do α in the objective world. Theorems 1c and 1d express the corresponding effects of the processes of intention generation and of intention reconsideration in the context of the agent's choices: after generating (resp. reconsidering) the intention to do α , the agent performs (resp. does not perform) action α in the context of his choices. Theorems 1e and 1f express that the operations of intention generation and of intention reconsideration are local operations, that is, the process of generating (resp. reconsidering) an intention does not affect the other intentions of the agent: if α and β are different actions and the agent intends (resp. does not intend) to do β then, after reconsidering (resp. generating) the intention do α , the agent will intend (resp. not intend) to do β .

Theorem 4. *The logic \mathbf{L}^{dyn} is completely axiomatized by principles in Fig. 1 together with the schemata of Theorem 3 and the rule of replacement of proved equivalence.*

Proof. By means of the principles **R1a-R6** in Theorem 3, it is straightforward to prove that for every \mathbf{L}^{dyn} formula there is an equivalent \mathbf{L} formula. In fact, each reduction axiom **R1a-R5c**, when applied from the left to the right by means of the rule of replacement of proved equivalence **R6**, yields a simpler formula, where 'simpler' roughly speaking means that the dynamic operators are pushed inwards. Once the dynamic operators attain an atom they are eliminated by one of the equivalences **R1a-R1c**. Hence, the completeness of \mathbf{L}^{dyn} is a straightforward consequence of Theorem 1. \square

4 Applications

It is now time to study in detail the notion of executability preconditions introduced in Section 3.2. We will study four general kinds of executability preconditions: the executability preconditions of the event which consists in making true (resp. false) a certain objective fact p ; the executability preconditions of an intentional action; the executability preconditions of a process of intention generation; the executability preconditions

of a process of intention reconsideration. Thus, we will be able to clarify why the distinction between atoms denoting facts and atoms denoting actions given in Section 2.1 is not merely a syntactic distinction. On the contrary, it has concrete effects on the dynamic level of our model. Indeed, we will show that the executability preconditions for assignments to atoms denoting facts are qualitatively different from the executability preconditions for assignments to atoms denoting actions.

Positive and negative effect preconditions. Executability preconditions can be used to describe how the world will change after the occurrence of a certain action, that is, how a fact p might become true (resp. false) when the agent acts in a certain way. In way similar to [15], we denote with $\gamma^+(\alpha, p)$ the positive effect preconditions of action α with respect to p (i.e. the conditions which ensure that p will be settled to be true when action α occurs) and with $\gamma^-(\alpha, p)$ the negative effect preconditions of action α with respect to p (i.e. the conditions which ensure that p will be settled to be false when action α occurs). For example, we might suppose that $\gamma^+(\text{pullTrigger}, \text{scaredEnemy}) = \text{holdsGun} \wedge \text{loadedGun}$, i.e. the positive effect preconditions of the action ‘pull the trigger of the gun’ with respect to the fact ‘the enemy gets scared’ consist in ‘holding a loaded gun in a hand’. As the following definition highlights, we can say that a certain fact will possibly become true (resp. false) if and only if, there exists an action α performed by the agent, the positive (resp. negative) effect preconditions of α with respect to p hold and there is no action β performed by the agent such that the negative (resp. positive) effect preconditions of β with respect to p hold.

Definition 1. For every $p \in ATM^{Fact}$ we define:

$$\begin{aligned} Pre(p \overset{W}{\rightsquigarrow} \top) &= \bigvee_{\alpha \in ATM^{Act}} (\alpha \wedge \gamma^+(\alpha, p)) \wedge \neg \bigvee_{\beta \in ATM^{Act}} (\beta \wedge \gamma^-(\alpha, p)); \\ Pre(p \overset{W}{\rightsquigarrow} \perp) &= \bigvee_{\alpha \in ATM^{Act}} (\alpha \wedge \gamma^-(\alpha, p)) \wedge \neg \bigvee_{\beta \in ATM^{Act}} (\beta \wedge \gamma^+(\alpha, p)). \end{aligned}$$

Executability preconditions for action execution. The following general principle clarifies the connection between mental level and intentional action level (a similar principle is discussed in [12]).

- (*) An action α will be possibly performed by an agent if and only if the agent has the intention to perform action α and he does not believe that doing α is something bad for him.

Therefore, if an agent intends to a certain action α and, before starting to perform the action, he learns that doing α is something bad for him, he will not start to execute the action and he will reconsider his corresponding intention (see Principle *** below). As the following definition highlights, the Principle * can be expressed in our logic.

Definition 2. For every $\alpha \in ATM^{Act}$ we define:⁵

$$Pre(\alpha \overset{W}{\rightsquigarrow} \top) = I(\alpha) \wedge \neg[B] \neg good_{\alpha}.$$

From Definition 2 it follows that all action occurrences of type α are occurrences of intentional actions.

⁵ We here suppose that the agent believes that doing an action α is *bad* for him if and only if he believes doing α is *not good* for him.

Executability preconditions for intention generation and for intention reconsideration. The executability preconditions of a process of intention generation correspond to general principles of instrumental rationality which specify the beliefs that an agent uses as premises of a practical argument (viz. the argument that concludes in an intention). Such beliefs are generally called *reasons for acting* or *reasons for intending* and have been extensively studied in the philosophical literature (see, e.g., [18, 1])

We here suppose that an agent will possibly decide to perform a certain action (and will possibly form the corresponding intention) on the basis of the following general principle of instrumental rationality:

(**) An agent will possibly form the intention to perform action α if and only if, he does not have this intention and he believes that doing action α is something good for him.

The Principle ** can be expressed in our logic in terms of the executability preconditions of the event $gen(\alpha)$.

Definition 3. For every $\alpha \in ATM^{Act}$ we define:

$$Pre(gen(\alpha)) = \neg I(\alpha) \wedge [B]good_{\alpha}.$$

The following principle is about intention reconsideration.

(***) An agent will possibly reconsider his intention to perform action α if and only if, he intends to perform action α and he believes that performing action α is something bad for him.

The Principle *** can be expressed in our logic in terms of the executability preconditions of the event $rec(\alpha)$.

Definition 4. For every $\alpha \in ATM^{Act}$ we define:

$$Pre(rec(\alpha)) = I(\alpha) \wedge [B]\neg good_{\alpha}.$$

An example We provide a general example in order to show how L^{dyn} can be concretely used to model intention dynamics. Consider the scenario represented in Fig. 2.

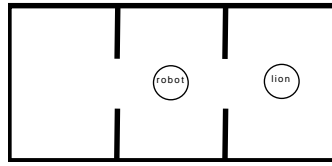


Fig. 2. Example

The agent is a robot moving in a space with three rooms (left room, middle room, right room). It can either move right or move left, that is, $ATM^{Act} = \{moveL, moveR\}$. Moreover, if the robot is in the middle room or in the left room and it moves to the left, it will be in the left room afterwards. If the robot is in the middle room or in the left

room or in the right room and it moves to the right, it will not be in the left room afterwards. Finally, if the robot is in the right room and it moves left, it will not be in the right room afterwards. These three facts are formally expressed by the following three positive and negative effect preconditions: $\gamma^+(moveL, robotL) = robotL \vee robotM$, $\gamma^-(moveL, robotL) = robotR$ and $\gamma^-(moveR, robotL) = robotL \vee robotM \vee robotR$.

The robot just entered into the middle room from the left room by moving right, it has still the intention to move right and it does not have the intention to move left in order to escape from the middle room:

H1. $robotM \wedge \neg robotL \wedge \neg robotR \wedge \mathbb{I}(moveR) \wedge \neg \mathbb{I}(moveL)$.

From the middle room the robot can see that a ferocious lion is inside the right room. Hence, the robot has the following beliefs: it believes that moving right is something bad for him and it believes that moving left is something good for him. Indeed, if it moves right it will be eaten by the lion, and if it moves left it will escape from the lion:

H2. $[\mathbb{B}]\neg good_{moveR} \wedge [\mathbb{B}]good_{moveL}$.

As the following proposition highlights, the previous Hypothesis H1 and H2 ensure that the following three-event sequence will possibly occur: the robot will reconsider its intention to move right and will generate the intention to move left; the robot will perform the action of moving left; the robot will enter into the left room. Moreover, at the end of the three-event sequence, the robot will be in the left room.

Proposition 2. $\vdash_{L^{dyn}} (H1 \wedge H2) \rightarrow \langle \{\emptyset, \{rec(moveR), gen(moveL)\}, \emptyset\}:w \rangle \langle \{\emptyset, \emptyset, \{moveL \xrightarrow{W} \top\}\}:w \rangle \langle \{\emptyset, \emptyset, \{robotL \xrightarrow{W} \top\}\}:w \rangle robotL$.

5 Related works and perspectives

Note that the three operators $[\Sigma:w]$, $[\Sigma:B]$ and $[\Sigma:C]$ can be seen as nothing but the three points e_W , e_B and e_C of an action/event model à la [7, 3], such that $Pre(e_W) = Pre(e_B) = Pre(e_C) = Pre(\Sigma)$ and such that $(e_W, e_B), (e_B, e_B), (e_C, e_C) \in \mathcal{B}$ and $(e_W, e_C), (e_B, e_C), (e_C, e_C) \in \mathcal{C}$, and such such that $Post(e_W)(p) = \sigma_W(p)$ for all atoms $p \in D(\sigma_W)$ and similarly for e_C and e_B , where $Post(e_W)(p)$ is the postcondition of the event e_W applied to the atom p . In the surprising variation on the event models with assignments proposed in this paper, we have defined preconditions *per atomic proposition* p and not all at once per event e_C , e_B , e_W . As only for a finite number of atoms such preconditions are given, the precondition à la DEL indeed corresponds to the conjunction defining $Pre(\Sigma)$ at the beginning of Section 4. So this is all neat and nice, and we have therefore more variation in finetuning preconditions than in standard DEL with assignments.

In [2, 11] a logic of knowledge and preference dynamics is provided. In van Benthem & Liu's approach knowledge dynamics are modeled by means of announcements (or updates), whereas preference dynamics are modeled by means of operations on accessibility relations called upgrades. We have provided here an approach to choice change and intention dynamics based on assignments. We think indeed that assignments, rather than announcements and upgrades, are more suited to model intention dynamics (intention generation and intention reconsideration). Indeed, intention dynamics

are obtained by means of *local* operations on an agent's choices and assignments are a natural candidate to formalize these kinds of operations. This locality aspect of intention dynamics has been discussed in Section 3.3 in which we have shown that a process of generating (resp. reconsidering) an intention defined in terms of assignments does not affect the other intentions of the agent (Theorems 1e and 1f).

Directions for future research are manifold. For instance, the logic \mathbf{L}^{dyn} does not allow to distinguish between *present-directed intention* and *future-directed intention*. According to [5], a future-directed intention is an intention to do some action later whereas a present-directed intention is an intention to do some action now. An interesting direction to be explored is an extension of \mathbf{L}^{dyn} with temporal modalities in order to be able to express this distinction. Furthermore, in this paper we only considered the single-agent case. We plan to extend our approach to the multi-agent case.

References

1. R. Audi. A theory of practical reasoning. *American Philosophical Quarterly*, 19:25–39, 1982.
2. J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17(2):157–182, 2007.
3. J. van Benthem, J. van Eijck, and B. Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, 2006.
4. P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.
5. M. Bratman. *Intentions, plans, and practical reason*. Harvard University Press, 1987.
6. P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.
7. H. van Ditmarsch and B. Kooi. Semantic results for ontic and epistemic change. In *Proceedings of LOFT 7*, pages 87–117, 2008.
8. H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Kluwer Academic Publishers, 2007.
9. J. Hintikka. *Knowledge and Belief*. Cornell University Press, 1962.
10. W. van der Hoek, W. Jamroga, and M. Wooldridge. Towards a theory of intention revision. *Synthese*, 155(2):265–290, 2007.
11. F. Liu. *Changing for the better: preference dynamics and agent diversity*. PhD thesis, University of Amsterdam, 2008.
12. E. Lorini and A. Herzig. A logic of intention and attempt. *Synthese*, 163(1):45–77, 2008.
13. J. J. Ch. Meyer, W. van der Hoek, and B. van Linder. A logical approach to the dynamics of commitments. *Artificial Intelligence*, 113(1-2):1–40, 1999.
14. A. S. Rao and M. P. Georgeff. Modelling rational agents within a BDI-architecture. In *Proceedings of KR'91*, pages 473–484, 1991.
15. R. Reiter. *Knowledge in action: logical foundations for specifying and implementing dynamical systems*. MIT Press, 2001.
16. Y. Shoham. Agent-oriented programming. *Artificial Intelligence*, 60:51–92, 1993.
17. M. Singh and N. Asher. A logic of intentions and beliefs. *Journal of Philosophical Logic*, 22:513–544, 1993.
18. G. H. Von Wright. On so-called practical inference. *The Philosophical Review*, 15:39–53, 1972.
19. M. Wooldridge. *Reasoning about rational agents*. MIT Press, 2000.